

Intelligibility of English Phonemes Spoken by L1 Japanese Speakers

Hideki Tani

Abstract

This study explored which segmental sounds of English spoken by Japanese L1 speakers were intelligible to non-native speakers of English. It also investigated the reasons for the reduced intelligibility of some of the sounds. Thus, this study addressed two research questions: RQ1 “What English phonemes spoken with a Japanese accent are intelligible enough in international settings?”, and RQ2, “If some phonemes are not intelligible, what are the reasons?”. Four Japanese speakers participated in this study and they read aloud 23 sentences which included 13 target sounds /a:, ʊ, u:, æ, ʌ, f, v, θ, ð, sɪ, ʃ, l, r/. Then, the intelligibility of the four speakers was measured by a Cloze Transcription Task in which 12 non-native listeners transcribed the utterances. The results indicated that the consonants were more intelligible than the vowels. However, the intelligibility of the consonants /θ/ and /l/ were shown to be as low as the vowels. An error analysis conducted in order to find the reasons of reduced intelligibility indicated some major findings. First, the vowels tended to get misrecognized as other sounds more often than the consonants. Second, /æ/ and /ʌ/ were often mistaken for each other. Third, the consonants /r/ and /l/ were indistinguishable to a large extent. Next, error analysis revealed that the intelligibility of the consonants in the final position influenced the recognition of words. Therefore, clear articulation of the consonants in the final position is recommended for Japanese users of English. Lastly, a large portion of the misrecognitions occurred when the words were monosyllabic. From these results, this study proposes implications for teaching which it is hoped will be applied in schools in Japan.

Keywords: intelligibility, phonemes, Japanese English

1. Introduction

In many parts of the world, English is widely spoken and it has now established its position as one of the most spoken languages in the world. Graddol (1997) showed that by 1997 the number of non-native speakers who used English for international communication outnumbered native speakers of English. Through this rapid growth of English, the pronunciation teaching paradigm has been changing.

Before the 1960s, the dominant paradigm in pronunciation teaching was based on the Nativeness Principle which assumed, “that it is both possible and desirable to achieve native-like pronunciation in a foreign language” (Levis, 2005, p. 370). The Nativeness Principle was the dominant paradigm and many people thought that native speakers were better than non-native speakers. However, Flege (1999) showed that having a foreign accent is unavoidable when learners start to learn the target language after the age of seven. Derwing and Munro (2005) also explain the attainment of the native-like pronunciation is difficult for foreign learners.

Considering the issues above, there is a contrastive view to the Nativeness Principle which is called the Intelligibility Principle (Smith, 2011). This principle is based on the idea that it is not desirable to aim for the acquisition of native-like pronunciation, but instead it is more desirable to acquire intelligible pronunciation because there are many models of English and learners may often communicate with non-native speakers of English. Furthermore, Smith and Nelson (1985) claim that native speakers are no longer the only judges of intelligibility. In the following quote, they refer to the issue of intelligibility:

Native speakers are not always more intelligible than non-native speakers. Given the same hearer, the speaker (native or non-native) who speaks clearly, is able to paraphrase, and talks at the appropriate level of the hearer in terms of proficiency, topic and speed will be most intelligible. (Smith & Nelson, 1985, p. 333)

With globalization, the opportunities to communicate with non-native speakers have been increasing. At the same time, acquiring intelligible pronunciation in order to communicate with non-native speakers has become even more important these days. For this reason, the kind of pronunciation that is intelligible enough for non-native speakers of English should be investigated.

There have been various intelligibility studies concerning many different English varieties, but few researchers have conducted studies about intelligibility of English in regard to Japanese English accents (hereafter JE) though some studies have investigated its phonological features or the difficult sounds (e.g., segmental features) for Japanese learners of English to pronounce (Y. Kachru & Nelson, 2006; Saito, 2011; Tsukada, 1996). Furthermore, even fewer studies have researched the intelligibility of JE to non-native listeners though there are a considerably greater number of

non-native speakers compared to native speakers. Therefore, this study investigates the intelligibility of the segmental features of JE for non-native speakers of English.

2. Literature Review

2.1 World Englishes

2.1.1 Background

World Englishes refers to the various distinct varieties of English spoken in the world. In the past English was used in the countries which were the colonies of Great Britain or the United States (e.g., Nigeria, India, the Philippines) as a means of communication among people who had different linguistic backgrounds. Recently, English has become widely used as a lingua franca (the language used among people who have different mother tongues) for social, economic, and political purposes in Europe, Africa, and Asia. From this history, many varieties of English have risen as local varieties of English (Goodwin, 2013).

2.1.2 Three Circles of English Varieties

B. Kachru (1992) made clear distinctions among three groups who use English. The first group is called “the inner circle”. The inner circle is the smallest group located at the center and the speakers in the inner circle are considered to be “native speakers” of English. The second group is “the outer circle”. It surrounds the inner circle and is the second biggest group. Like Indians or Singaporeans, speakers in the outer circle use English locally as their second language. The biggest group is “the expanding circle” which is located in the outermost ring. The expanding circle includes countries such as China, Japan, Russia, or Turkey which use English as a foreign language.

2.2 The English Variety Spoken by Japanese Native Speakers

2.2.1 Differences between Japanese and English

There are several important differences between Japanese and English. First of all, the total number of phonemes is different between Japanese and English. Above all, the numbers of vowels in these two languages widely differs. While Japanese has five vowels, /a, i, u, e, o/, American English has 12 vowels (Enomoto, 2000). In contrast, there is not much difference in the number of consonants between these two languages. It is said that Japanese has 24 consonants (/p, b, t, d, k, g, m, n, ɲ, ŋ, ɳ, ɾ, ɸ, s, z, ɕ, ʃ, h, ts, tɕ, ɕʃ, j, ɰ/) and English has also 24 consonants (/p, b, t, d, k, g, m, n, ŋ, l, w, r, f, v, θ, ð, s, ʃ, ʒ, z, ʃ, ʒ, h, j/) (Imaishi, 2005; Kubozono, 1998), but these consonants do not overlap completely.

Differences exist at the supra-segmental level as well. Japanese is an open syllable language and this means that most syllables end with a vowel except for /N/ as in ‘kan’ (can), or double consonants as in ‘hipparu’ (pull) (Kubozono, 1998). On the other hand, English is a closed syllable

language in which only 40% of the utterances in natural conversation end with a vowel. In addition, Japanese is a pitch accent language in which speakers emphasize in certain sounds by controlling the pitch whereas English is a stress accent language in which the emphasis depends on the intensity of voice (Kubozono, 1998).

2.2.2 Phonological Features of JE

One of the characteristics found in JE is the insertion of vowels to simplify consonant clusters because Japanese rarely have consonant clusters (Kubozono, 1998). The second feature of JE is the sound substitution or confusion of some particular consonants. For example, /θ/, as in ‘thing’ or ‘thrill’, is substituted with /s/. Similarly, Japanese speakers of English often substitute /z/ or /ʒ/ for /ð/. For example, ‘These’ sounds like /zi:z/. Furthermore, /l/ is often substituted with /r/, and /v/ is substituted with /b/. Confusion of these sounds are often found in the utterances from Japanese speakers of English (Y. Kachru & Nelson, 2006). Furthermore, the substitution between /s/ and /ʃ/ is also common among Japanese learners of English (Saito, 2011).

Tsukada (1996) investigated the difference of the frequency of the use of vowels between native speakers of English and Japanese speakers of English. There were six Japanese participants and five of them had studied in English speaking environments for at least six years. She investigated the difference of formants 1 and 2 using acoustic analysis (A formant is defined as “the spectral peaks of the sound spectrum” [Fant, 1960, p. 20]. These two formants mainly determine the vowel quality). She found Japanese speakers of English tend to confuse the sounds of /æ/ and /ʌ/. Also the /u:/ and /u/ sounds were confused.

Apart from the segmental sounds, further differences are found in supra-segmental features. Taniguchi (2001) claims that the pitch movement of Japanese speakers is much narrower than that of the native speakers and it is suggested that this narrow pitch movement affects listeners’ comprehension of the listeners.

2.3 Lingua Franca Core

For a long time, teachers have been wondering what model to teach when they give pronunciation instruction. Jenkins (2002) found that many factors affected intelligibility in international communication in her previous studies, and so, she created a pronunciation syllabus for non-native speakers of English who use English as an international language. The syllabus is called the Lingua Franca Core (LFC) and it is shown in Table 1.

LFC includes how some sound features affect the speakers’ intelligibility and provides an order of priority for pronunciation teaching. In the table, the right most column shown as the “EIL target” is essential in this study, and teachers are recommended to refer to the sound features in the

far right column. For example, it is shown that all consonants except for /θ/, /ð/, and /ʌ/ influence intelligibility, so it is recommended to teach these three consonants after all the others.

Table 1

Pronunciation Syllabus from Jenkins (2002)

		NS target	EIL target
1.	The consonantal inventory	<ul style="list-style-type: none"> ● all sounds ● RP non-rhotic /r/ ● GA rhotic /r/ ● RP intervocalic [t] ● GA intervocalic [ɾ] 	<ul style="list-style-type: none"> ● all sounds except for /θ/, /ð/, /ʌ/ ● rhotic /r/ only ● intervocalic [t] only
2.	Phonetic requirements	<ul style="list-style-type: none"> ● rarely specified 	<ul style="list-style-type: none"> ● aspiration after /p/, /t/, /k/ ● appropriate vowel length before fortis/ lenis consonants
3.	Consonant clusters	<ul style="list-style-type: none"> ● all word positions 	<ul style="list-style-type: none"> ● word initially, word medially.
4.	Vowels quantity	<ul style="list-style-type: none"> ● long-short contrast 	<ul style="list-style-type: none"> ● long-short contrast
5.	Vowel quality	<ul style="list-style-type: none"> ● close to RP or GA 	<ul style="list-style-type: none"> ● L2 (consistent) regional qualities
6.	Weak forms	<ul style="list-style-type: none"> ● essential 	<ul style="list-style-type: none"> ● unhelpful to intelligibility
7.	Features of connected speech	<ul style="list-style-type: none"> ● all 	<ul style="list-style-type: none"> ● inconsequential or unhelpful
8.	Stress-timed rhythm	<ul style="list-style-type: none"> ● important 	<ul style="list-style-type: none"> ● does not exist
9.	Word stress	<ul style="list-style-type: none"> ● critical 	<ul style="list-style-type: none"> ● unteachable/can reduce flexibility
10.	Pitch movement	<ul style="list-style-type: none"> ● essential for indicating attitudes and grammar 	<ul style="list-style-type: none"> ● unteachable/ incorrectly linked to NS attitudes/grammar
11.	Nuclear (tonic) stress	<ul style="list-style-type: none"> ● important 	<ul style="list-style-type: none"> ● critical

Note. NS = Native Speaker

EIL = English as an International Language

2.4 Intelligibility

2.4.1 Definition

To gain a better understanding of intelligibility, a definition of intelligibility is provided here. Munro and Derwing (1999) give a broad definition of intelligibility as “the extent to which a speaker’s message is actually understood by a listener.” (Munro & Derwing, 1999, p. 289). Smith and Nelson’s definitions (1985) of intelligibility and comprehensibility are slightly different from

those in Munro and Derwing (1999) and Munro and Derwing (1995a) although they are similar. According to Smith and Nelson (1985), the notion of intelligibility is limited to the recognition of words or utterance.

2.4.2 Measuring Intelligibility

There are various ways to measure speech intelligibility (Munro & Derwing, 1999). One of the methods is the Word Recognition Task. The Word Recognition Task is a task in which listeners transcribe words one by one. The advantage of this task is that the contextual or grammatical cues can be removed. However, Miller (2013) suggests that the Word Transcription Tasks are not natural for real conversations because many people are “not restricted to single words” (Miller, 2013, p. 604) in real life conversations.

Another way common method used in intelligibility studies is a form dictation task in which listeners hear not single word utterances but sentences and write them out in standard orthography (e.g., in Derwing & Munro; 1997, Munro, Derwing, & Morton, 2006). A third well-known way is the sentence cloze test in which listeners transcribe some parts of the utterances (Smith & Rafiqzad, 1979).

2.4.3 Variables in Intelligibility Studies

Zielinski (2006) investigated how non-native speech from Vietnamese speakers affected intelligibility. According to the results, the listeners misunderstood the intended words when one consonant was substituted for another consonant, a consonant was added, or a consonant was missing. A large number of misunderstandings occurred when consonants were mispronounced in the final position. It was implied that listeners rely on non-standard segmental speech signals such as the sound position more than non-standard syllable stress patterns.

However, Zielinski (2008) also conducted a study of English spoken by L1 Korean, Mandarin, and Vietnamese speakers and investigated their intelligibility to English native listeners. It was revealed that stresses in inappropriate positions can often lead to the lack of intelligibility. Field (2005) concluded that native speakers of English strongly rely on the positions of stressed sounds and non-standard word stresses may negatively influence intelligibility.

The difference between vowels and consonants influence intelligibility, too. In the study from Cole et al. (1996), it was investigated to what extent vowels and consonants contribute to word intelligibility using sentences in which all the vowels were deleted by noise, all the consonants were deleted by noise, and unmodified sentences in which nothing is deleted. The results revealed that the vowels contributed to the recognition of words more than consonants.

2.5 Research Questions

As Graddol (1997) showed, the rapid increase in the number of non-native speakers of English has changed the role of English communication but there have been few studies which focus on the non-native listeners. Jenkins (2002) found that many segmental sounds had an influence on intelligibility and Saito (2011) investigated the difficult phonemes for Japanese learners of English to pronounce. Thus, it is essential to research the contributing and non-contributing phonemes which affect intelligibility. Therefore, this study addresses the following two research questions:

RQ1: What English phonemes spoken with a Japanese accent are intelligible enough in international settings?

RQ2 : If some phonemes are not intelligible, what are the reasons?

3. Method

3.1 Participants

The speakers were four male native speakers of Japanese at a private university in West Japan aged 22 to 23. They had Test of English for International Communication (TOEIC) scores ranging from 472 to 550. Although they had some opportunities to use English in their academic lives, they only used Japanese in their daily lives. They had had no experience of studying abroad and they all had received similar English education (three years in junior high school, three years in high school, and two to three years at university). They had not had any special English education or pronunciation training.

The listeners were 12 non-native speakers of English (2 Germans, 1 Italian, 1 Mexican, 1 Indonesian, 1 Malaysian, 5 Chinese). The reasons for including only non-native listeners were that there have been many intelligibility studies which have used native listeners, but there have been fewer studies using non-native listeners. There were 11 women and one man. They all had had experience of living in English speaking environments. Their ages ranged from 23 years old to 30 years old. The length of residence in Japan varied from one month to two years. The listeners did not have any common source to indicate their English level, so they were asked to self-evaluate their English proficiency. The self-evaluation sheet was based on the six-level can-do list of Common European Framework of Reference for Languages (Council of Europe, 2001). The listeners' English proficiencies varied from B2 to C2.

3.2 Stimulus Sentences

In order to investigate the first research question, the stimulus sentences included the sounds which were considered to be difficult for Japanese speakers or believed to affect intelligibility. In this study, the sounds /ɑ:, ʊ, u:, æ, ʌ, f, v, θ, ð, sɪ, ʃ, l, r/ were chosen because of the following

reasons.

The vowels /ɑ:/, ʊ, u:/, æ, ʌ/ were chosen in reference to Tsukada's study (1996). Tsukada (1996) shows that Japanese speakers tend to confuse vowels when speaking in English. Especially, /ʊ/ and /u:/, or /æ/ and /ʌ/ are often indistinguishable. Y. Kachru and Nelson (2006) show that the /v/ is often substituted with /b/ and /l/ is often substituted with /r/. Furthermore, /θ/ and /ð/ are often confused with other sounds. Although /sɪ/ is a syllable not a phoneme, it is often confused with the /ʃ/ sound by Japanese learners when /t/ comes after it (Saito, 2011), so it was included in the stimulus sentences. These sounds are shown to be important for foreign learners of English to distinguish in Brown (1988). Brown (1988) ranks the substitutions between /æ/ and /ʌ/, /r/ and /l/ as rank ten. There are ten ranks in Brown (1988) so rank ten is the highest rank and substitutions in rank ten are the most important as well as the most frequent errors made by the foreign learners. The common confusions among Japanese learners such as /b/ and /v/, /s/ and /ʃ/, /ð/ and /z/, and /h/ and /f/ are also regarded as highly important distinctions (above rank seven). In contrast, the common substitution /θ/ and /s/ and /ʊ/ and /u:/ are not regarded as important distinctions (below rank five).

After the target phonemes were determined, the stimulus sentences were chosen. The sentences were picked from the book "*Voice and Articulation Drillbook*" (Fairbanks, 1960). This book is for speech language hearing therapists and there are exercise sentences which cover every phoneme in English. From the book, 21 sentences were chosen. As for the /sɪ/ sound, it was not included in the book. Thus, two stimulus sentences for this sound were created. In total, 23 sentences were used as stimulus sentences (Appendix).

3.3 Task Used in This Study

Intelligibility was measured by a Cloze Transcription Task which had between four and ten word blanks in each sentence. There were two reasons of adapting a cloze task. The first reason was to reduce the burden on the listeners. Since the stimulus sentences were not made for an intelligibility study, it was considered that the listeners would not be able to concentrate on transcribing all the words in the sentences. The second reason was to eliminate vocabulary issues for the listeners. It was considered that the listeners' vocabulary sizes varied, so none of the words of Level 6 and above in JACET 8000 (Aizawa, Ishikawa, Murata et al., 2005) were transcribed (Level 5 is considered to be university level.).

One more variable may be the contextual or grammatical cues given by the speakers may help the listeners guess what has been said and the present author was aware of it. However, as Miller (2013) suggests Word Recognition Tasks which eliminate the contextual and grammatical cues by having only one-by-one word transcriptions are not appropriate for natural conversation and many researchers have adapted the transcription tasks using sentences instead (Derwing & Munro, 1997; Munro & Derwing, 1995a; Rooy, 2009; Smith & Rafiqzad, 1979). Therefore, in this study the

transcription was done in sentences. Since there are blanks in each sentence and the aim of the study is to identify intelligibility of the target phonemes, the number of each target phoneme was balanced. Each target phoneme appears more than four times.

3.4 Procedure

3.4.1 Recording Sessions

Each recording session took place in a quiet room in the university. When the list of the 23 stimulus sentences was handed to each speaker, they were asked whether there were any words that they had not seen before. The reason for doing this was to reduce the risk of the speakers not being able to pronounce some words because of their lack of familiarity with them. The speakers' English proficiency was maximum TOEIC 550, so it was expected that there would be some words that the speakers had not seen before. When there were such words, the speakers were allowed to either ask the researcher how to pronounce the words, or check in a dictionary. Each sentence was recorded with an OLYMPUS "Voice Trek V-822". When the speaker made a mistake, she was asked to read the same sentence aloud again.

3.4.2 Cloze Transcription Task

In the Cloze Transcription Task session, each listener listened to the stimulus sentences from two speakers. In short, each listener transcribed two sets of 23 sentences (46 sentences in total). The order of the listening is shown in Table 2. Three out of 12 listeners listened to the same speakers in the same order. For example, as Table 2 shows Listeners 1, 2, and 3 transcribed the 23 sentences from Speaker 1 first, then they listened to the same 23 sentences from Speaker 2. In contrast, Listener 4, 5, and 6 listened to the utterances from Speaker 2 first, and transcribed the utterances from Speaker 2 as the second set.

First, the task was explained. Each listener understood that even if some words were unintelligible that was not a problem because the aim of the study was to identify which phonemes affect intelligibility. Therefore, it was necessary that the listeners transcribe the utterances exactly as they heard the sentences, and in order for the listener to be able to do so, some instructions on how to transcribe were given in advance. The instructions were given using the English orthography rules written in the *Macmillan Dictionary* (William, 1979). To practice, the listeners listened to non-existent words (Rastle, Harrington & Coltheart, 2002) and transcribed the words using the English orthography rules.

During the Cloze Transcription Task each listener confirmed whether she had finished transcribing each sentence in order to give enough time to complete transcribing. After the first set was finished, the listeners were given a five- to ten-minute break. In the second set, the listeners listened to the same set of sentences. When the task was finished, the answers were handed to the

listeners. At the very end of the whole procedure, the listeners self-evaluated their English proficiency using the CEFR can-do statement. They were given enough time to evaluate their English proficiency and they were allowed to leave when they finished the self-evaluation.

Table 2

Grouping of Listeners

Listeners	Groups	Order of Listening
Listener 1 Listener 2 Listener 3	Group 1	Speaker 1 → Speaker 2
Listener 4 Listener 5 Listener 6	Group 2	Speaker 2 → Speaker 1
Listener 7 Listener 8 Listener 9	Group 3	Speaker 3 → Speaker 4
Listener 10 Listener 11 Listener 12	Group 4	Speaker 4 → Speaker 3

3.5 Analysis

In this study, the intelligibility scores were not calculated as in many previous studies which quantified how many words were transcribed out of the total number of words in the utterances. Since the aim of this study is to investigate the intelligibility of phonemes, only the target phonemes were quantified. For example, /ʊ/ appeared four times in each set of the task, so each listener transcribed /ʊ/ eight times (four times in the first set and four times in the second set). 12 listeners did the same task and the total number of tokens of /ʊ/ was 96 ($8 \times 12 = 96$). Then, the 96 tokens of /ʊ/ were calculated and converted into percentages (what percent of /ʊ/ was correctly transcribed out of 96). /ʊ/ and /ð/ were the sounds which appeared the least since they were transcribed 96 times. /ɪ/ was transcribed the most at 240 times.

As for RQ2, “If some phonemes are not intelligible, what are the reasons?”, an error analysis was conducted. The error analysis was conducted in terms of the following four criteria:

1. What sound is each target sound mistaken and mistranscribed as?
2. Where do mistranscriptions of the consonants occur? In other words, do they occur in the initial, middle, or final position?
3. Does the number of syllables of the words affect vowel intelligibility?

4. Is reduced intelligibility led by the incorrect placement of stress?

As for Criterion 1, “What sound is each target sound mistaken and mistranscribed as?”, the /l/ sound was categorized into two types of sounds. The first one is light /l/ as in the words such as ‘light’ or ‘clear’ and the second one is dark /l/ (/ɫ/) as in the words such as ‘feel’ or ‘milk’ because Jenkins (2002) found that the dark /l/ tend not to influence intelligibility.

The reason why Criterion 2, “Where do mistranscriptions of the consonants occur? In other words, do they occur in the initial, middle, or final position?” focuses on only consonants is due to the task selection. Most of the target vowels from the stimulus sentences occurred in the middle position of the words. Therefore, it was decided that the vowels were irrelevant to this issue.

Criterion 3, “Does the number of syllables of the words affect vowel intelligibility?” was added because there was a variation in the number of syllables among the words containing the target vowels.

Criterion 4, “Is reduced intelligibility led by the incorrect placement of stress?” was included because JE is said to have non-standard stress (Taniguchi, 2001). Furthermore, the incorrect placement of stress tends to reduce intelligibility (Field, 2005; Zielinski, 2008).

4. Results

4.1 Results of the Cloze Transcription Task

Table 3 shows the descriptive statistics of the intelligibility scores of the target phonemes from the four speakers. The numbers in the parenthesis after each target sound indicates the total tokens of the target sound transcribed by the listeners. For example, the number “120” after the sound /sɪ/ indicates that /sɪ/ was transcribed 120 times in total by the 12 listeners (five times in the stimulus sentences × two sets × 12 listeners = 120 in total). As stated in the procedure and the analysis sections, each listener transcribed two sets of the 23 stimulus sentences from two speakers. The sounds which were transcribed the least were /ʊ, ɔ/ (96 times), and the sound which was transcribed the most was /r/ (240) times.

When we look at the individual scores of the speakers individually, the scores ranged from 52% to 100%. The mean scores of the target sounds of the four speakers were calculated and they ranged from 64% to 94%. The sound /sɪ/ appeared to be the most intelligible having an intelligibility score of 94% whereas /l/ was the least intelligible. Only 64% of the /l/ sound was transcribed correctly by the listeners. All the consonants except for /θ/ and /l/ were more intelligible than vowels. Furthermore, among the vowels the long vowels (e.g., /u:/, /a:/) were more intelligible than short vowels.

Table 3

Intelligibility Score (Ratio of Correctly Recognized Phonemes)

	Speaker 1	Speaker 2	Speaker 3	Speaker 4	mean
si (120) ^a	90%	93%	97%	97%	94%
ð (96)	91%	100%	96%	83%	93%
f (144)	89%	89%	97%	94%	92%
r (240)	92%	97%	90%	75%	89%
v (120)	83%	90%	83%	97%	88%
ʃ (144)	94%	92%	75%	89%	88%
u: (144)	100%	97%	67%	58%	81%
a: (168)	81%	83%	83%	71%	80%
ʌ (120)	87%	80%	80%	70%	79%
ʊ (96)	88%	96%	58%	63%	76%
θ (120)	80%	73%	70%	63%	72%
æ(120)	60%	67%	57%	83%	67%
l (168)	64%	67%	52%	73%	64%

Note. ^a Number of the total tokens of the specific sounds.

4.2 Results of the Error Analysis

In order to find the reasons for reduced intelligibility, an error analysis was conducted. The analysis was done following the four criteria listed in the analysis section. However, it should be reported here that as for the fourth criterion, “Is reduced intelligibility led by the incorrect placement of stress?”, it was checked if there were any words pronounced with non-standard stress. Only one error was found from the recording from the four speakers. When Speaker 4 pronounced ‘understand’, he pronounced /ʌnd'ə:stænd / placing tonic stress on the second syllable. Except for this, none of the words were pronounced with non-standard stress, so it was decided not to include Criterion 4 “Is the reduced intelligibility led by the change in stress?” any further.

Table 4 shows the sounds each target sound was mistaken for and transcribed as. It also shows their sound positions when the listeners did not transcribe anything, it is indicated as “t.d.” (Total Disregard). For better understanding of the table, some explanation of the table is necessary. The farthest left column shows the target phonemes in this study. The number in the parenthesis right below each target phoneme shows the total number of the mistranscriptions of the target sound. For example, the number (23/96) after /ʊ/ shows that this sound was mistranscribed 23 times out of the 96 total transcriptions. It has to be noted here that the number of the total mistranscriptions (23 in the case of /ʊ/) does not include t.d. which is located in the farthest right column (As for /ʊ/ and /a:/, it is located at the furthest right column in the second row). The reason for not including t.d. in the number of the total mistranscriptions is that the calculating methods were different between

mistranscriptions and Total Disregards. In the case of /ʊ/, the total transcriptions were 96, and the mistranscriptions were 23. Percentages in t.d. were calculated from the number of total transcriptions (96 tokens) and the percentages in mistranscriptions were calculated from the number of the total mistranscriptions (23 tokens). For example, the percentage 20% shown in the “→/u:/” column of target /ʊ/ means that 20% of 23 total tokens of mistranscription were mistranscribed as /u:/. Therefore, when the percentages of the mistranscription of each target sound (every number except for the number in t.d.) are added together, it becomes 100% (error range = ± 2%). For t.d., the ratio was calculated per total number of each target sound transcribed. In the case of /ʊ/, this sound was disregarded four times out of 96 times. Therefore, the scores for the total disregard of /ʊ/ is 4% ($4 \div 96 = 0.041$).

The sounds after the symbol “→” represent the mistranscribed sounds of each target sound. Therefore, for example, the target sound /ʊ/ was mistranscribed to /u:, e:, i, ʌ, ue, o, e, o:/. Mistranscriptions in each sound position are shown using the symbol ‘○○○’ in the second column from the left. This shows whether the mistranscription occurred in the initial, middle, or final position’. When the first ‘○’ is underlined, (○○○), this means that the mistranscription of the particular phoneme occurred when it was located in the initial position of a word. When the second ‘○’ is underlined (○○○), the target sound in the middle position was mistranscribed. Finally, when the last ‘○’ is underlined (○○○), this means that the mistranscription occurred in the final position of the word. The ratio (%) of the mistranscription which occurred in each sound position was shown using the number in the parenthesis. For example, /ʊ/ in the middle position was mistranscribed as /u:/ with the probability of 20% out of the total 23 mistranscriptions. Some of the sound positions show “n.a.”. This stands for “not applicable” and indicates none of the words containing the target sound had that particular position. For example, as for the /ʊ/ sound, there was no word in the stimulus sentences which started or ended with this sound. Therefore, the initial position and the final position were not applicable and this was indicated as “n.a.”.

Table 4 shows that there are more frequencies of errors in vowels (e.g., 23/96 for /ʊ/). Among the vowels, there was significant confusion between /æ/ and /ʌ/. 71% (14% in the initial position + 57% in the middle position) of the total mistranscriptions of /ʌ/ (26 tokens) were shown to be mistranscribed as /æ/ (18 tokens), and 60% of the total mistranscriptions of /æ/ (39 tokens) were mistranscribed as /ʌ/ (23 tokens).

Table 4 shows that 80% of the total mistranscriptions of the light /l/ (34 tokens) were mistranscribed as /r/ (27 tokens). Furthermore, 64% of /r/ (17 tokens) were mistranscribed as /l/ (11 tokens). Therefore, it is considered that there was significant confusion between /r/ and /l/. As for /h/, the table shows this sound was never mistaken for another sound. Rather, this sound has a tendency to be missed by listeners (23% of the total transcriptions of /h/ (96 tokens) were totally disregarded by the listeners).

Table 4

Mistranscribed Sounds and Error Positions Found from the Error Analysis

	Tokens of mistranscription in each position		→/u:/	→/e:/	→/i:/	→/Λ/ (→/ue/
/ʊ/ (23/96) ^a	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
	□□□ (23)	(20%)	(20%)	(20%)	(13%)	(7%)	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
		→/o/	→/e/	→/o:/	t.d.		
	□□□	(n.a.)	(n.a.)	(n.a.)	(n.a.)		
	□□□	(7%)	(7%)	(7%)	(4%)		
/æ/ (39/120)	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
	□□□ (39)	(60%)	(30%)	(7%)	(3%)	(15%)	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
/Λ/ (26/120)	Tokens of mistranscription in each position	→/æ/	→/e	→/ei/	→/a:/	t.d.	
	□□□ (5)	(14%)	(5%)	(0%)	(0%)	(1%)	
	□□□ (21)	(57%)	(10%)	(10%)	(5%)	(1%)	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
/u:/ (28/144)	Tokens of mistranscription in each position	→/il/	→ /i/	→/e:/	→/ʊ/	t.d.	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
	□□□ (28)	(64%)	(21%)	(7%)	(7%)	(0%)	
	□□□ (na)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
/a:/ (34/168)	Tokens of mistranscription in each position	→/o:/	→/æ/	→/au	→/o/	→/i/	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
	□□□ (34)	(30%)	(17%)	(13%)	(9%)	(9%)	
	□□□ (0)	(0%)	(0%)	(0%)	(0%)	(0%)	
		→/ie:/	→/ai/	→/e/	→/ei/	→/Λ/	t.d.
	□□□	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
	□□□	(4%)	(4%)	(4%)	(4%)	(4%)	(6%)
	□□□	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)

/f/ (12/144)	Tokens of mistranscription in each position	→ /t/	→ /v/	→ /s/	t.d.	
	□□□ (0)	(0%)	(0%)	(0%)	(0%)	
	□□□ (2)	(0%)	(14%)	(0%)	(0%)	
	□□□ (10)	(57%)	(14%)	(14%)	(3%)	
/v/ (14/120)	Tokens of mistranscription in each position	→ /l/	→ /f/	→ /d/	→ /b/	t.d.
	□□□ (na)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)
	□□□ (4)	(14%)	(0%)	(0%)	(14%)	(0%)
	□□□ (10)	(29%)	(29%)	(14%)	(0%)	(1%)
/θ/ (44/120)	Tokens of mistranscription in each position	→ /s/	→ /z/	→ /f/	t.d.	
	□□□ (22)	(38%)	(9%)	(3%)	(0%)	
	□□□ (0)	(0%)	(0%)	(0%)	(0%)	
	□□□ (22)	(50%)	(0%)	(0%)	(1%)	
/ð/ (7/96)	Tokens of mistranscription in each position	→ /z/	→ /d/	→ /f/	→ /k/	t.d.
	□□□ (0)	(0%)	(0%)	(0%)	(0%)	(0%)
	□□□ (0)	(0%)	(0%)	(0%)	(0%)	(0%)
	□□□ (7)	(33%)	(33%)	(17%)	(17%)	(0%)
/ʃ/ (16/144)	Tokens of mistranscription in each position	→ /s/	→ /f/	→ /z/	t.d.	
	□□□ (0)	(0%)	(0%)	(0%)	(3%)	
	□□□ (16)	(76%)	(13%)	(13%)	(1%)	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	
/l/ (34/72)	Tokens of mistranscription in each position	→ /r/	→ /d/	→ /f/	t.d.	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	
	□□□ (34)	(80%)	(17%)	(4%)	(8%)	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	
/ʌ/ (0/96)	Tokens of mistranscription in each position	t.d.				
	□□□	(n.a.)				
	□□□	(19%)				
	□□□	(4%)				

/r/ (17/240)	Tokens of mistranscription in each position		→/l/	→/d/	→/v/	→/w/	t.d.
	□□□ (5)	(0%)	(23%)	(0%)	(7%)	(3%)	
	□□□ (12)	(64%)	(0%)	(7%)	(0%)	(1%)	
	□□□ (n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	(n.a.)	
/s/ / (2/125)	Tokens of mistranscription in each position		→/so/	t.d.			
	□□□ (2)	(100%)	(3%)				
	□□□ (0)	(0%)	(1%)				
	□□□ (n.a.)	(n.a.)	(n.a.)				

Note. ^a Number of the mistranscription out of the total transcription

□□□ = indicator of the sound position of the mistranscriptions

n.a.= not applicable.

t.d. = total disregard

An observation of the ratio of mistranscription of the consonants over the three different positions appears to have a certain tendency, i.e., frequent mistranscription in the final position. Thus the details were analyzed focusing on this issue. In order to make a fair comparison, it is necessary to check whether the test sentences had even distribution of the target sounds over the three positions. It has to be noted that only /f, v, θ, ð/ were relevant to this issue because there were no words which ended with /ʃ, l, r/ in the stimulus sentences. Though there were words which end with /t/, as stated above /t/ was not mistranscribed as other sounds. Therefore, it was decided not to include this sound in this issue. /f/ appeared six times in the stimulus sentences. Two were in the initial position (e.g., ‘father’, ‘food’), two were in the middle position (e.g., ‘coffee’, ‘infants’), and two were in the final position (e.g., ‘staff’, ‘life’). Similarly, there were six words with /θ/ in the stimulus sentences, and /θ/ appears twice in the initial position (e.g., ‘thought’, ‘theory’), twice in the middle (e.g., ‘pathetic’, ‘something’), and twice in the final position (e.g., ‘faith’, ‘truth’). Although the number of sound positions was balanced for /f/ and /θ/, it was not completely balanced for /v/ and /ð/. /ð/ appeared four times. One was in the initial position (e.g., ‘this’), two were in the middle position (e.g., ‘father’, ‘weather’), and one was at the final (e.g., ‘breathe’). /v/ had no initial position. Among five tokens of /v/ in the stimulus sentences, three were in the middle position (e.g., ‘heavy’, ‘revived’, ‘eventually’), and two were in the final position (e.g., ‘believe’, ‘save’).

As shown above, the number of the words appearing in each sound position is balanced for /f, v, θ, ð/ to some extent and it is assumed that the number of mistranscriptions could have been divided equally (the mistranscriptions in the final position should occur with 20% for /f, θ/, 40% for /v/, and 25% for /ð/) if the frequency of mistranscription had occurred regardless of the position of

the sounds. However, Table 4 indicates 100% (7 tokens) of mistranscribed /ð/ (7 tokens), 85% (10 tokens) of mistranscribed /f/ (12 tokens), 72 % (10 tokens) of mistranscribed /v/ (14 tokens), and 50% (22 tokens) of mistranscribed /θ/ (44 tokens) occurred in the final position. These ratios are considerably greater than the assumed ratios. Figure 1 shows concisely that a large portion of mistranscriptions of /f, v, θ, ð/ occurred in the final position (the part highlighted in black is the ratio of mistranscriptions).

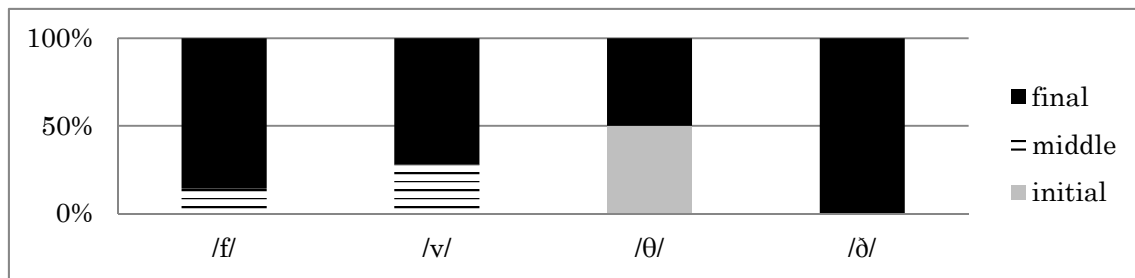


Figure 1 .The Ratio of the Position of the Mistranscription of /f, v, θ, ð/

Note. /v/ did not appear in the initial position of words.

As for, Criterion 3, “Does the number of syllables of the words affect vowel intelligibility?”, the results are presented in Table 5. It shows that the number of syllables does influence intelligibility as intelligibility tended to be lower with the monosyllabic words. Table 5 shows that except for /ʊ/, every other target vowel had much lower intelligibility when the word carried only one syllable (/ʊ/ = 81%, /æ/ = 55%, /ʌ/ = 26%, /u:/ = 53%, /a:/ = 62%). However, intelligibility increases when the words carry more than two syllables. For these vowels, intelligibility of the bisyllabic words containing the target vowels ranged from 75% to 96% while intelligibility of the trisyllabic words ranged from 90% to 96%. Therefore, it is much more difficult to recognize the sound of the monosyllabic correctly than the bi- or tri-syllabic words.

Table 5

Mean Intelligibility Scores of the Vowel Sounds and the Number of Syllables in Words

	Monosyllabic	Bisyllabic	Trisyllabic
/ʊ/	81%	75%	n.a.
/æ/	55%	83%	n.a.
/ʌ/	26%	92%	96%
/u:/	53%	96%	90%
/a:/	62%	94%	n.a.

Note. n.a. = not applicable

5. Discussion

5.1 Gaps in Intelligibility Among the Target Sounds

Some of the sounds were less intelligible than the other sounds even though the target sounds are all regarded as difficult sounds for Japanese learners of English to pronounce (Tsukada, 1996; Lambacher et al., 2005; Y. Kachru & Nelson, 2006; Saito, 2011). The overall lower intelligibility of vowels can be explained using Tsukada's study (1996). According to the acoustic analysis conducted by Tsukada (1996), some of the English vowels spoken by Japanese learners were confused (e.g., /æ/ and /ʌ/, /ʊ/ and /u:/). Therefore, it is implied that Japanese speakers of English tend to pronounce these vowels identically. In this study, the confusion or the substitution of these vowels confused the listeners and may have lowered intelligibility.

Gaps in intelligibility were found among the target consonants. The sounds /l/ and /θ/ had lower intelligibility than the other consonants. A possible explanation for the low intelligibility of /l/ is that as Y. Kachru and Nelson (2006) claim, Japanese speakers tend to substitute the /l/ sound with /r/. Brown (1988) says that distinction between /r/ and /l/ is one of the most important phoneme distinctions and also the most frequent errors among foreign learners. This may be the case with the Japanese speakers in this study.

The phoneme /θ/ is usually considered to be a sound which does not affect intelligibility (Graddol, 1997; Jenkins, 2002). However, in this study this sound was shown as the second least intelligible consonant. A possible explanation for this is due to the task instrument used in this study. The 14th stimulus sentence was, "But we had *faith that something would lead to the truth*". The listeners were supposed to transcribe the part in italics (see also Appendix). More than 45% of the /θ/ mistranscriptions occurred with the word 'faith' in the sentence being mistranscribed as 'faced'. In this study, this was the only mistranscription replaced by one specific word with this high ratio. It is considered that the low intelligibility of /θ/ can be explained by this task instrument issue. Therefore, we need further research to conclude whether /θ/ is a particularly difficult sound to recognize.

As for the sounds which had relatively high intelligibility (/si, ð, f, r, v, ʃ/), it can be implied that these sounds are intelligible enough to non-native listeners although they have been said to be among the difficult sounds for Japanese learners in the previous literature (Brown, 1988; Y. Kachru & Nelson, 2006; Lambacher et al., 2005; Saito, 2011; Tsukada, 1996). Therefore, when teaching pronunciation, it is suggested that paying too much attention to these phonemes (/si, ð, f, r, v, ʃ/) may not be as fruitful and necessary as focusing on the vowels and some important consonants such as /l/ and /r/.

5.2 Major Findings from the Error Analysis

An error analysis was conducted in order to answer RQ2 using the criteria shown in the analysis section. There were five major findings: (1) mistranscription of vowels occurred more than

that of consonants, (2) confusion of /æ/ and /ʌ/, (3) confusion of /r/ and /l/, (4) a large proportion of mistranscribed consonants occurred in the final position, (5) a high ratio of monosyllabic words were mistranscribed and disregarded. In this section, the five findings are going to be discussed and compared with previous studies in order to find similarities and differences.

5.2.1 Mistranscription of Vowels Occurred More than That of Consonants

The study from Tsukada (1996) can be again used to explain some of the reasons for this finding. As stated above, Japanese speakers of English tend to confuse certain vowels. This is reinforced in this study because a greater number of mistranscribed vowels than consonants was found. The mean number of the kinds of mistranscribed vowels was greater than that of consonants ($M = 6.0$ for the vowels, 3.1 for the consonants). This means that the speakers may have given the listeners redundant choices when recognizing words, and reduced vowel intelligibility. In other words, confusion concerning pronunciation can also give confusion to listeners and may lower intelligibility. Therefore, attaining clear distinctions between sounds, especially of vowels, is strongly recommended for foreign learners.

The results of this study showed some differences from previous studies in relation to this finding. In the *Lingua Franca Core* from Jenkins (2002), it is recommended to make a clear distinction between phonetically similar long and short vowels such as /i/ and /i:/ because it affects intelligibility. This study included two short and long vowels (/ʊ/ and /u:/) which are phonetically similar because it was expected that these two phonemes would be confused mutually hence reducing intelligibility as Tsukada (1996) shows. The results revealed that only 20% of the tokens of /ʊ/ were mistranscribed as /u:/, and 7% of the tokens of /u:/ were transcribed as /ʊ/. Therefore, similar long and short vowels were not mutually confused by the listeners in this study to a large extent. However, this does not mean that it is not necessary to teach the distinction between short and long vowels, so it is still important to be able to distinguish the difference between each vowel because the confusion of vowels may have had a negative influence on the listeners' perception hence intelligibility.

5.2.2 Confusion of /æ/ and /ʌ/, /r/ and /l/

The results of the error analysis indicated that the 71% of the tokens of /ʌ/ were mistranscribed to /æ/ and 60% of the tokens of /æ/ were mistranscribed as /ʌ/. 80% of the tokens of /l/ were mistaken and transcribed to /r/ and 62% of the tokens of /r/ were mistranscribed as /l/. For Japanese learners of English, distinguishing between the /r/ and /l/ sounds is difficult because these two sounds do not exist separately in the Japanese language. Tsukada (1996) claims distinguishing /æ/ and /ʌ/ sounds for Japanese speakers is also difficult and many of them tend to confuse these two. As noted earlier, Y. Kachru and Nelson (2006) say there is a tendency of substitutions between /l/ and /r/

sounds among Japanese learners. According to Brown (1988), the distinction of these sounds is one of the most important distinctions for foreign learners. The substitutions among these sounds were found in high proportion in this study, so the teaching of these sounds is strongly recommended.

As for /ʌ/, Jenkins (2002) reported that it did not affect intelligibility. However, it seemed to be a sound that influenced intelligibility in this study. This sound was the only sound which was not mistaken as any other sound, but many listeners missed this sound completely (23% of the tokens of this sound were totally disregarded). This means that /ʌ/ was often not recognized by the listeners. Therefore, as well as the light l (/l/), the teaching of the allophone of the /l/ sound (/ʌ/) is recommended.

5.2.3 A Large Proportion of Mistranscribed Consonants Occurred in the Final Position

The results also show that mistranscriptions occurred many times when the consonants /f, v, θ, ð/ were located in the final position of words. A possible reason for this is a phonological feature of the Japanese language. In Japanese, it is rare that consonants appear in the final position of words because Japanese is an open syllable language in which 90% of words end with a vowel (Kubozono, 1998). Therefore, it is reasonable that Japanese speakers find it difficult to pronounce consonants in the final position of words whereas they find it relatively easier to pronounce the consonants in the initial position because they do pronounce consonants in the initial position in their L1.

This finding is in line with the findings of Zielinski (2006). According to Zielinski (2006), reduced intelligibility occurred the most due to lack of clearly pronounced consonants in the final position and the least in the initial position. Zielinski (2006) also claimed, listeners strongly rely on the consonant in the final position of a word when recognizing words. Therefore, rather than paying attention to the consonant in the initial position, paying attention to the consonant in the final position is much more helpful for gaining intelligibility.

Among the mistranscriptions in the final position, there were some unpredicted recognitions. For example, /f/ was mistranscribed to /t/ ('staff' → 'stunt', 'stert') and /s/ ('life' → 'nice'). /ð/ was mistranscribed to /ʃ/ ('breathe' → 'rish') and /k/ ('breathe' → 'break'). /v/ was mistranscribed to /l/ ('save' → 'seal') and /d/ ('revive' → 'devide'). These mistranscriptions were not supposed to occur because they are neither the substitutions of the target sounds nor phonetically similar sounds. However, if the listeners are relying on the speech signals in the final position of each word, some explanation can be found. The listeners who participated in this study may have not been able to hear the chopped off, or substituted sounds in the final position (e.g., Speaker 1 pronounced 'staff' as /stʌʔ/). This might have led the listeners to end up guessing words. It is important to note that these unpredicted mistranscriptions did not occur with any particular speaker, but they occurred to all the speakers who participated in the study.

5.2.4 A High Ratio of Monosyllabic Words were Mistranscribed and Disregarded

Except for /ʊ/, a large portion of the mistranscriptions occurred with monosyllabic words. Generally speaking, recognizing bi- or poly-syllabic words is easier than recognizing monosyllabic words because bi- or poly-syllabic words have greater redundancy. In other words, the longer the word is, the more clues the listeners can use to guess what the word is. Previous studies such as Cole (1996) and Fogerty and Humes (2010) parallel this result as they show that vowels carry more contextual cues and hence contribute to intelligibility in sentences. Polysyllabic words obviously carry more vowels and this may have enabled the listeners to guess the words. Therefore, foreign learners have to pay less attention to polysyllabic words for they will be understood anyway, but they should pay more attention to the clear and intelligible pronunciation of monosyllabic words.

6. Conclusion

In this study, the intelligibility of English segmental sounds uttered by Japanese native speakers of English was investigated in order to answer RQ1, “What English phonemes spoken with a Japanese accent are intelligible enough in international settings?”. Intelligibility was measured by a Cloze Transcription Task in which 12 non-native listeners transcribed stimulus sentences which included 13 target sounds from four Japanese speakers. The results revealed the gaps in intelligibility between the vowels and the consonants. Briefly, most consonants except for /l/ and /θ/ were found to be more intelligible than the vowels. An error analysis conducted to answer RQ2, “If some phonemes are not intelligible, what are the reasons?” showed that the vowel or consonant qualities, the sound positions, and the number of syllables influenced intelligibility. Vowels were mistranscribed as many other sounds, while consonants were not transcribed to as many sounds as the vowels. The sound position affected the consonant intelligibility as intelligibility was lowered when consonants at the final position of words were pronounced. Furthermore, it was revealed that intelligibility was lowered when the listeners listened to monosyllabic words. Some of the findings may be limited to only JE (e.g., /r/ and /l/ confusion), but others may be generalized to other non-standard varieties of English. For example, articulating clearly in the final position of words may be recommended for foreign learners whose L1s are open syllable language (e.g., Italian).

However, there are limitations to this study. First of all, one limitation is that there were both grammatical and contextual cues when the listeners were taking the Cloze Transcription Task since they listened to sentences, not words respectively. There are several reasons why the transcription was done in sentences. As noted earlier, Miller (2013) claims transcribing only word for word is unnatural because usually people are not restricted to listening to single words in their daily lives. Thus, in many intelligibility studies the transcriptions have been done in sentences (Bent and Bradlow, 2003; Derwing and Munro, 1997; Rooy, 2009; Smith and Rafiqzad, 1979). On a daily basis, we use our contextual and grammatical knowledge to understand utterances and this was considered

to be a natural thing to do in this study. This was also the reason for using sentences in the transcription task. If the transcriptions had been done word for word in this study, the grammatical and contextual clues would have been reduced, but at the same time the task would not have been the one which simulated the authentic conversations. However, the present author hopes to find a way to solve this dilemma as he wishes to continue this study.

Another limitation was specific to this particular study. As stated in the discussion section, there was an issue affecting intelligibility due to the task instrument. The /θ/ sound had low intelligibility and this can partly be explained due to the selection one of the stimulus sentences. Therefore, if a further study is to be conducted along these lines, attention will need to be paid to the selection of the stimulus sentences. Only one pilot study was carried out in this study, so more pilot studies would be desirable.

Lastly, the implications for teaching will be discussed here. Table 3 shows intelligibility of each target sound and it shows the priority of phoneme teaching. It is desirable that the sounds with low intelligibility such as /l/ or vowels are taught with priority. The error analysis revealed many findings, too. In short, five findings have important implications for the pronunciation teaching:

1. Clear distinctions between English vowels
2. Clear distinction of /æ/ and /ʌ/
3. Clear distinction of /r/ and /l/
4. Clear articulation of the final consonants of each word
5. Clear articulation of monosyllabic words rather than bi- or poly-syllabic words

As for first one, clear distinctions should be made between each vowel sound because Japanese learners tend to confuse vowels when they pronounce them and these confusions do affect intelligibility. In particular, the confusion of /æ/ and /ʌ/ had a significant negative influence. Therefore, these two vowels should be carefully taught. The consonants were not as confused as the vowels, but /r/ and /l/ were confused by the listeners. A clear distinction between /r/ and /l/ is necessary. Although it seems to be difficult for Japanese learners to pronounce consonants in the final position, it is recommended to learn how to pronounce consonants clearly especially in the final position to increase speech intelligibility. In addition, it is important to clearly pronounce monosyllabic words rather than bi- or poly- syllabic words because listeners have few clues to help them understand monosyllabic words.

Hopefully, some of those findings will be used to develop students' intelligibility and this study will be of help of the English education in Japan.

Reference

- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 114(3), 1600-1610. doi:10.1121/1.1603234
- Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, 22(4), 593-606.
- Cole, R. A., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. *International Conference on Acoustics, Speech, and Signal Processing*, 2, 853-856 doi:10.1109/ICASSP.1996.543255
- Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1-16.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379-398.
- Enomoto, M. (2000). *Phonetics of English and Japanese speech sounds*. Tokyo: Tamagawa Daigaku Shuppan.
- Fairbanks, G. (1960). *Voice and articulation drillbook*. New York: Harper and Row.
- Fant, G. (1960). *Acoustic theory of speech production: Description & analysis of contemporary standard Russian*. DE: Mouton De Gruyter.
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399-424.
- Flege, J.E. (1999). Age of learning and second-language speech. In D.P. Birdsong (ed.), *Second language acquisition and the critical period hypothesis* (p. 101-132). Hillsdale, NJ: Lawrence Erlbaum.
- Goodwin, J. (2013). Teaching Pronunciation. In Celce-Murcia, M, Brinton, D & Snow, M (Eds.), *Teaching English as a second or foreign language*. (p. 136- 152). Boston: National Geographic Learning.
- Graddol, D., & English 2000 (Project). (1997). *The future of English?: A guide to forecasting the popularity of English in the 21st century*. London: British Council.
- Imaishi, M. (2005). *Onsei gaku Kenkyu Nyumon*. [An introduction to phonetics]. Osaka: Izumi Shoin.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83-103.
- Kachru, B. B. (1992). Models for non-native Englishes' B.B. Kachru (ed): *The Other Tongue. English Across Cultures*. Urbana, IL: University of Illinois Press.
- Kachru, Y., & Nelson, C. L. (2006). *World Englishes in Asian contexts*. Hong Kong: Hong Kong

University Press.

- Kubozono, H. (1998). *Nichiei taisho ni yoru eigo gakushu series 1. Onsei gaku Oninron*. [Comparison between Japanese and English. English learning series 1. Phonetics Phonology]. Tokyo: Kuroshio Shuppan.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369-378.
- Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, 48(6), 601-612. doi:10.1111/1460-6984.12061
- Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(s1), 285-310. doi:10.1111/0023-8333.49.s1.8
- Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38 (Pt 3)(3), 289-306.
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520-531. doi:10.1016/j.system.2006.09.004
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111-131. doi:10.1017/S0272263106060049
- Rooy, S. C. (2009). Intelligibility and perceptions of English proficiency. *World Englishes*, 28(1), 15-34. doi:10.1111/j.1467-971X.2008.01567.x
- Saito, K. (2011). Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English. *RELC Journal*, 42(3), 363-378. doi:10.1177/0033688211420275
- Smith, J. (2011). Teaching pronunciation with multiple models. *New Zealand Studies in Applied Linguistics*, 17(2), 107-115.
- Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes*, 4(3), 333-342. doi:10.1111/j.1467-971X.1985.tb00423.x
- Smith, L. E., & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, 13(3), 371-380.
- Taniguchi, M. (2001, April). *Japanese EFL learners' weak points in English intonation*. Paper presented at the 2001 Phonetics Teaching and Learning Conference, London. Retrieved from <http://www.phon.ucl.ac.uk/home/johnm/ptlc2001/pdf/tani2.pdf>
- Tsukada, K. (1996). Acoustic analysis of Japanese-accented vowels in English. In P. McCormick & A. Russell (Eds.), *Proceedings of the 6th Australian International Conference on Speech Science and Technology* (pp. 373-378). Canberra: Australian Speech Science and Technology

Association.

William D. Halsey (Ed.) (1979). *Macmillan Dictionary*. New York: Macmillan Publishing.

Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. *Prospect*, 21, 22-45.

Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36(1), 69-84. doi:10.1016/j.system.2007.11.004

Appendix

1. 以下の 24 つの文を音読してください

*読み方がわからない単語は事前に聞いてください

1	Captain Hook pushed through the bushes to the brook.
2	From where he stood it looked like an ambush.
3	Hugh refused to join the musician's union.
4	His excuse was viewed with amusement.
5	Sally banged the black sedan into a taxicab.
6	It was badly damaged by the crash.
7	I am unable to understand my uncle Gus.
8	He mutters and mumbles about nothing.
9	"For breakfast", said father, coffee is the staff of life.
10	Grapefruit is a food for infants.
11	I believe I'll save this heavy veil.
12	The vogue might be revived eventually.
13	We thought that the theory was pathetic.
14	But we had faith that something would lead to the truth.
15	My father finds it hard to breathe in this weather.
16	The fishing ship was in shallows near the shore.
17	In one motion a wave crushed it on the shoal.
18	Lawyer Clark held his felt hat and his gloves in his lap.
19	He silently placed the will on the table.
20	John started across the yard toward the barn.
21	His father remarked calmly that he'd better not wander too far.
22	You can see good seafood in this season cooked and served in various ways all over the city.
23	The secret of the taste is the spices used in each restaurant.